



A Requirement-driven Approach for Designing Data Warehouses

Camille Salinesi, Ines Gam

► To cite this version:

Camille Salinesi, Ines Gam. A Requirement-driven Approach for Designing Data Warehouses. Requirements Engineering: Foundations for Software Quality (REFSQ'06), Jun 2006, Luxembourg, Luxembourg. pp.1. halshs-00176276

HAL Id: halshs-00176276

<https://shs.hal.science/halshs-00176276>

Submitted on 3 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Requirement-driven Approach for Designing Data Warehouses

Ines Gam, Camille Salinesi

Centre de Recherche en Informatique
Université Paris 1 - Panthéon Sorbonne
90, rue de Tolbiac, 75013 Paris, France
Ines.Gam@univ-paris1.fr; Camille.Salinesi@univ-paris1.fr

Abstract. A data warehouse (DW) is an integrated and historised collection of data generally used to make strategic decisions by means of online analytical processing techniques. Most of the existing DW development tools used nowadays in the industry focuses on the structures for data storage, e.g. applying the star or snowflake schema. We believe that DW that better suit the needs of decision makers would be delivered by concentrating more on their requirements. So far, very few approaches have been proposed to elicit DW requirements. This paper proposes a method, called *CADWA*, for guiding DW analyst in the elicitation of decisions-makers requirements and in their operationalization into a DW model. *CADWA* shifts the focus from where information from on how they should be structured and why they are needed. To comply with current practice, the approach starts with the elicitation of high-level requirements, reuses a set of data mart (DM) models and produces a model for the new DW. The paper presents each stage of the *CADWA* approach, and provides illustrations with an example inspired from a real case.

1. Introduction

Warehouses (DW) have become a standard tool used in many organizations to support decision making. Most of the existing DW development approaches deal with how data should be structured, stored, and managed in DW. Most often, the effort is concentrated on reconciling structural and semantic differences into a uniform data model. These approaches deal with conceptual data models, the logical data models and the physical data models, but they do not explicitly take requirements into account. This is confirmed by studies from the Standish Group that show that a 1/3 of DW projects fail to meet their objectives. The main problems are: (i) poor communication between IT people and decision makers, and (ii) poor project management. As stated by (Schiefer and al., 2002), “a great deal of RE effort and planning is required to achieve successful DW implementation”. Very few approaches consider requirements in advance by asking the question of why the DW is needed (e.g. (Freudi and al., 2003), (Prakash and al., 2004) and (Màzon and al. 2005)). This is

probably due to the time constraints imposed to DW projects and general belief that requirements-driven approaches are time consuming.

However, based on our experience, we believe that DW that better suit the decision-makers expectations would be delivered by focusing on the capture of their requirements. The *CADWA* approach bridges that gap by taking into account both decision-makers' requirements, operational data models of existing systems, and repositories of reusable Data Marts (DM)¹ models. The guidance offered by *CADWA* helps eliciting early requirements, refining them, and concretizing them into a logical data DW model. We think that putting this method into practice can, indeed, be time-consuming, but once this preparation step has been carefully achieved, it can be expected that higher quality results will be produced when a decision-maker asks for new information to support a decision. Our expectation is also that overall, having anticipated the query will help provide answers more quickly.

The rest of this paper is structured as follows. Section 2 presents an overview of the proposed process and its application on the example of a business retailer's DW requirements analysis. Section 3 discusses related works. Section 4 presents our conclusions based on results obtained after a first validation step achieved by discussing the pros and cons of our approach with consultants in the decision domain.

2. Overview of *CADWA*

There are several differences between the goals of transactional Information System (IS) and a DW system; our position is that the following differences advocate for adapting traditional general purpose requirements engineering methods, techniques and tools to the specific context of DW systems.

- First, traditional IS are expected to implement business rules. This is not the case for DW systems, which rather concentrate on data gathering.
- Second, in a traditional IS, data usually follow a CRUD lifecycle. Therefore, their value change over time and they can be deleted. In a DW, all data values must be preserved through historisation to support analysis of past situations.
- Another aspect of IS data vs. DW data is their origin. Whereas most IS data are directly derived from human activity (through user transactions), DW data are mostly extracted from other systems.
- The interfaces provided by IS to support transactions are usually designed once, and then stay stable, whereas the interfaces of DW systems change each time a new analysis is requested by the decision-makers.
- Last, traditional IS users have an operational activity, whereas DW users are decision makers. This is a determining difference as the expectation of traditional IS users is that the IS supports them in achieving their business process whereas the decision makers' goal is to monitor such processes.

Two times phases can be distinguished in the analysis of a DW system: (i) the initial DW development project (once before its first use), and (ii) the additional

¹ A DM is a small DW with a restricted scope of content and support for analytical processing, serving a particular data analysis problem domain.

development phases needed on specific parts of the DW when an analysis is requested by a decision maker (several times while the system is in use). Our position is that because of the time constraints it might be difficult to achieve a requirements engineering method at phase (ii)². On the contrary, time is available during phase (i) to anticipate decision makers' analysis requests. *CADWA* proposes to elicit requirements by anticipating these requests. The other sources of information taken into account in *CADWA* are the enterprise strategic objectives, decision-makers' objectives, the structure of existing transactional IS, existing DW models and reusable generic DM models.

Fig. 1 presents the process proposed to guide the *CADWA* method. As the figure shows it, the main goal is to produce a new DW model, the approach being to exploit requirements. The process is composed of three phases which purpose is respectively to: (i) elicit requirements, this is done using a goal based approach (ii) design a DW model fragment, this is done by selecting DM based on modelling indications associated with goals and (iii) integrate all the DW model fragments into a global DW model.

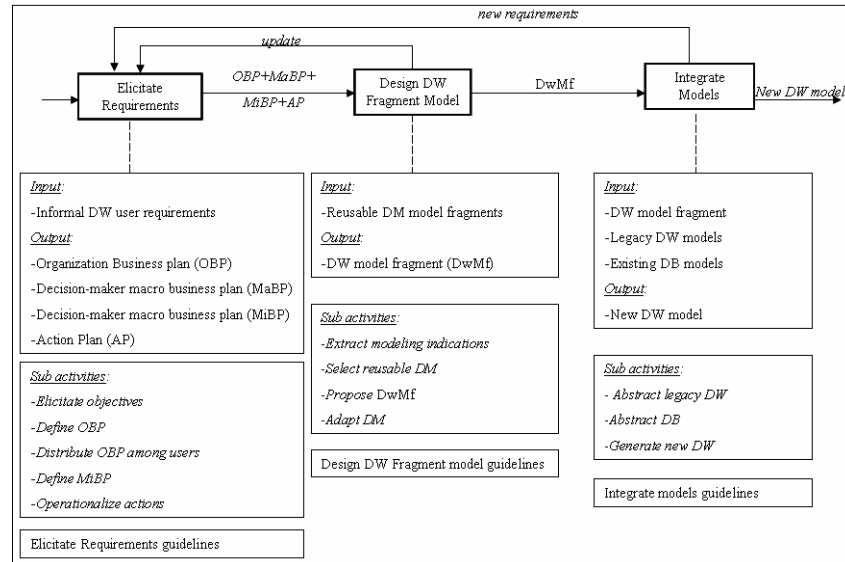


Fig. 1. Overview of the *CADWA* Process

The example chosen to illustrate *CADWA* is that of a business retailer. This example was chosen because of the diversity of requirements involved and of the nature of the analysis to achieve the “maximize profits” objective. This example was developed based on several business retailers' cases studies. We were also greatly inspired by documentation issued by the first retailer in Europe (and second largest world wide³). The retailer owns shops in four main formats (hypermarket,

² In our experience, decision-makers are usually executives who expect results in an hour time frame, and managers who expect results in a few days rather than in a few months.

³ According to 2004 statistics.

supermarket, hard discount and convenience store) over 32 countries. Each of the 868 hypermarkets offers to the customer a mix of food and non-food products together with different services (e.g. travel insurances, sales services, bookings and parking). 2.376 supermarkets sell about 10.000 items in a convenient layout so that shopping can be done without losing time and assuring the quality of fresh products, quality of customer relations and competitive prices. Each of the 4,934 hard discounts shops offers a selected range of food products, mostly retailer branded in order to offer low prices. Finally, convenience stores and other business stores are district or village shops. They present a range of products covering food requirements, as well as a range of services such as Cash & Carry, catering outlets, or electronic commerce.

The rest of this section details the guidance provided at each stage of the process, then illustrates application of the stage with the example.

2.1 Elicitation of requirements

The first stage of the *CADWA* process deals with early requirements gathering. The initial approach is to study informal needs expressed by decision-makers. Top level requirements are summarized into the first strategic plan of the organization in order to provide an understanding of how business owners, managers, or entrepreneurs intend to organize entrepreneurial endeavour and implement activities necessary and sufficient for the venture to succeed. This is consistent with (Prakash and al., 2003), who states that the analysis of DW requirements must aim at analyzing, understanding, and modeling the context of a DW.

Strategic plans are captured by goals organised in an *organization business plan*. While the organization business plan concerns the whole organization, its different parts involve different decision-makers that have different expectations with respect to the DW system (as second-level users). Therefore, consistent collections of requirements by regrouping goals from the organization business plan depending on the interest and activity of each decision maker.

Experience in real DW projects shows that it is not easy to distribute requirements among users. Indeed, different users with different activities can have different view on the same requirement, for instance on the way to refine to-level goal into more detailed ones. To deal with this issue, requirements are organized into four levels: *Organization business plan*, *Decision-maker macro business plan*, *Decision-maker micro business plan* and *Action plan*. The following paragraphs give more precisions on these different levels.

2.2 Designing requirements with Map

The goal-based approach to requirements modelling was chosen to obviate the difficulty with large amount of details and multiple points of views. From the existing goal models (i.e. MAP, KAOS, I^{*}, or even Use Case models) we chose especially the Map formalism. The choice of Map is motivated by its ability to represent multiple concerns at once, and by the fact that it helps making a clear distinction between what to achieve (intentions) and the different ways to achieve it (strategies). Our experience with business process modelling, change management and ERP installation and (Ben

achour and al., 2000), standardisation and globalization of practices across branches of a company, and integration due to company merge/take-over showed that Map is qualified for abstracting requirements by focusing on organizational goals and on their achievement.

Map provides a representation mechanism based on a flexible ordering of intentions and strategies. A map is a labelled directed graph with *intentions* (goals to achieve or maintain) as nodes, and *strategies* (means or manners to attain a goal) as edges. Map models are composed of *sections* which are triplets $\langle I_i, I_j, S_{ij} \rangle$ where I_i is the source intention, I_j the target intention and S_{ij} the strategy to attain when I_i has been achieved. The directed nature of the graph shows which intentions can follow which one. An edge enters a node if its strategy can be used to achieve the target intention of the node. Since, there can be multiple edges entering a node, a map can describe the different strategies to achieve an intention.

2.3 Organization Business Plan

An *organization business plan* (OBP) identifies the business objectives and business opportunities of the decisions makers, and represents them at the top level of abstraction in the DW requirements chain. These objectives stand as the primary benefits that the DW system will provide to the organization and its users. They are thus considered as early DW requirements. As Fig1 shows it, this is achieved in CADWA by reformulating the *informal DW user requirements* into an *organization business plan* (cf. “Define OBP” activity). Many techniques and tools, such as PEST or Balanced Score Cards, are available on the markets to identify business objectives. Guiding their usage is out of scope of CADWA. Our assumption is that these goals are given or documented (e.g. in an annual report, financial documents or in shareholders information). The main purpose of CADWA is to guide the re-formulation and structuration of these high level objectives into a goal model that can be used as a starting point to identify the DW requirements.

In the business retailer example (BR), the first needs of the BR directors are to adapt the retailer’s organization to its size and complexity and to be more in line with its structures and demands. The indicators provided by the management information system⁴ show that the BR should develop especially its external growth. Therefore, the directors decide to make progress in four important directions: *Adjust price position*, especially in the hypermarkets in France, *Improve growth and profitability of international businesses*, *Rationalize business portfolio* and *Strengthen financial position*. These four important directions are described as intentions and can be used to compose the organization business plan. For the sake of space, we focus in the rest of the paper on two particular intentions: (b) *Adjust price position* and (c) *Improve profitability of international business*. The corresponding part in the organization business plan is detailed in Fig. 2.

As the Map in Fig. 2. shows it, each intention is achieved with several strategies. Intention (b) can be achieved by two strategies (1) *By reducing cost product* and (2) *By increasing the number of customers*. Two sections are thus defined in the organization business plan, namely ab1 and ab2. Besides, the directors aim at

⁴ Management information systems are used to make business intelligence. They provide information on the business environment, such as market trends, competitors, legislation, etc.

achieving the intention (c) by three different strategies (1) *By reducing the net indebtedness*, (2) *By selling non strategic and non profitable business* and (3) *By improvement of financial ratios*. Consequently, the organization business plan is composed of three supplementary sections: ac1, ac2 and ac3.

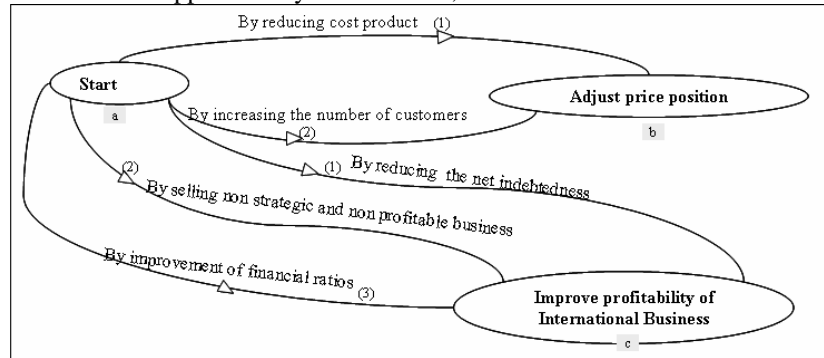


Fig. 2. Part of retailer organization business plan's

2.4 Decision-makers' Macro Business Plans

The purpose of the second step of requirement elicitation is to identify *decision-maker macro business plans* (MaBP). Contrary to organization business plans, which are global plans defined at the level of the organization, decision-maker macro business plans identify goals at the local level of decision makers. Decision-maker macro business plans are defined using a distribution matrix.

DW user requirements must comply with the context established by the decisions makers. Therefore, the combination of all decision-maker macro business plans must be consistent with the global organization business plan. CADWA defines decision-maker macro business plans according to their activities. This is done by distributing the sections of BP among users according to their activities (cf. "distribute OBP among users" activity). Each group of users, represented by a decision-maker, has the responsibility (i) to achieve its own part of the BP (ii) to make sure that its plan is not inconsistent with the other decision-maker macro business plans and (iii) to make sure that all decision-maker macro business plans contribute to the higher-level strategic goals.

As the Table 1 shows it, two particulars decision-makers can be considered in the BR example: the Marketing Manager and Financial Manager.

Decision maker \ BP section	ab1	ab2	ac1	ac2	ac3
Marketing Manager	*	*			
Financial Manager			*	*	*

Table1. Distribution matrix defining two MaBP in the BR example

According to the American Marketing Association, marketing is the process of planning and executing the pricing, promotion, and distribution of goods, ideas, and services to create exchanges that satisfy individual and organizational goals. In this

view, a marketing manager MaBP is a part of the proposed organization business plan that details the actions necessary to achieve specified marketing objectives. It can be for a product or service, a brand, or a product line, and it can cover one year (referred to as an annual marketing plan), or cover up to many years.

In our case, the marketing manager MaBP is restricted to the part of the organization business plan that corresponds to sections ab1 <Start, Adjust price position, By reducing cost product> and ab2 <Start, Adjust price position, By increasing the number of customers>. Besides defining the affectations of the responsibility of these two sections of the organization business plan to the Marketing Manager, the Marketing Manager MaBP offers a first entry into the collections of requirements that the Marketing Manager has with respect to the DW system. Indeed, it identifies the context in which the Marketing Manager will use the system: the decisions to make to achieve these goals. A more detailed view of these requirements is provided by the Decision-maker micro business plan.

2.5 Decision-maker Micro Business Plan

Decision-maker micro business plan (MiBP) documents the operational requirements for the DW system. For each section of a decision-maker micro business plan, there is a collection of actions that can be undertaken. CADWA guides the identification of these actions using refinement mechanisms (cf. “Define MiBP” activity). Each decision-maker micro business plan can be described into more details with a single map that contains lower-level intentions and strategies. Each section in a decision-maker micro business plan is refined by a path (i.e. a subset of sections) in the corresponding decision-maker micro business plan.

In the BR example, refining section *ab2*: <Start, adjust price position, by increasing the number of customers> allows entering into the details of the pricing plan which belongs to the more global strategy of the Marketing Manager.

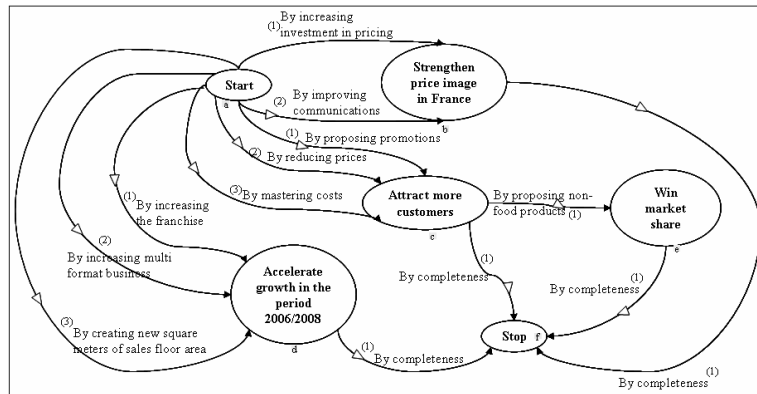


Fig. 3. Marketing Manager micro business plan

The detailed map of the Marketing Manager’s MiBP is based on intentions (cf. Fig. 3): (b) *Strengthen the price image in France*, (c) *Attract more customers*, (d) *Accelerate growth in the period 2006/2008* and (e) *win market share*. The Marketing

Manager has decomposed his/her plan into thirteen sections (*ab1*, *ab2*, *ac1*, *ac2*, *ac3*, *ad1*, *ad2*, *ad3*, *bf1*, *ce1*, *ef1*, *df1* and *ef1*). The Marketing Manager can choose between three sections to attract more customers: *ac1*, *ac2* and *ac3*. Section *ac3* aims at analyzing how to use a low cost product to attract consumers. Once the retailer has established a relationship with consumers via the low cost product, he/she sells additional higher-margin products and services. The principle underlying *ac1* is the same, but on a reduced time span; the prices of a set of products are temporarily reduced. On the contrary, the price reduction in *ac2* can be longer, if not definitive. The path (*ac1*, *ac2*, *ac3*) defines the Marketing Manager MiBP plan in refinement of the *ab1* section of the overall Organization business plan “adjust price position by reducing products costs”.

2.6 Action Plan

The business plans specified so far are high level views of candidate business objectives. These objectives are not, per se, functional requirements for the DW system. However, the knowledge needed to monitor their achievement or to decide between them can be considered as high level functional requirements for the DW system. There is thus a direct relationship between these business plans and the DW early requirements, which could be specified as a system goal: “monitor <section>”. These requirements are however high level and more details are required. Detailed requirements can be grasped by entering into the details of the *action plan* needed to put the candidate business plans into practice. Decision makers want to monitor action plans to evaluate their decisions or make new ones. Therefore, their requirements on the DW are to access information on each part of their action plans.

CADWA uses (Prat, 1997)’s linguistic approach to express actions as intentions in action plans (cf. “Operationalize actions” activity). Action intentions are structured as follows:

Verb<Target> [<Parameter>]*

where different mandatory and optional parameters can be used depending on the verb class. Different types of parameters, such as time, location, object, result, source or destination, define the semantic role that parameters can play with respect to the action intention. Characterizing each parameter for a given action intention helps generalizing the instance-level parameters so as to identify the information required to monitor actions once put into practice, or to measure the hope of success of an action intention.

In the BR example, the role of the Marketing Manager is to determine whether a promotion is effective or not in the context of the business plan section *ac1*: < *Start, Attract more customers, by proposing promotions*>. In this context, the Marketing Manager deals with analyzing and measuring each of the following action intentions:

AI1: Sell _(verb) 40% more _(quantity) childcare products _(object) in French hypermarkets _(location) during the Christmas period _(time).

AI2: Reduce _(verb) products returns _(object) in hypermarkets _(location) after the promotion periods _(time).

AI3: Reduce _(verb) products returns _(object) from market segment 4 _(source) to stores _(destination) on after the promotion periods _(time).

AI4: Sell _(verb) 20.000 items _(quantity) of product SKU D-042-0000073465-3 _(object) in convenience stores _(location) on weekend days _(time) with promotional conditions _(means).

AI5: Push _(verb) the gross sales _(result) of baby wipes _(object) with promotions on <wipe+diaper> packages _(means).

AI6: Propose _(verb) special sales _(result) to families _(beneficiary).

Generalizing the parameters of action intention AI leads to the conclusion that information is required on: (i) gross sales quantities (quantity parameter), (ii) products (object parameter), (iii) stores (location parameter), and (iii) time period (time parameter). Each of these action intention parameter is likely be analyzed by the Marketing Manager. Designing a data model to store these required information is guided by CADWA in the next phase of the process.

3. Designing DW fragment model

3.1 Mapping heuristics

CADWA proposes to guide the DW design based on a collection of heuristic rules. The *Map to Star* heuristic rules ($M2^{*}_{Rules}$), help extracting indicators in order to determine some of the dimensions and facts to be added to the DW model. These rules indicate for example that an object or a result of an action intention can be proposed as a fact or a fact table. Direction parameters of action intentions (source, destination) can be proposed as a dimension or an attribute of a dimension, etc. The choice of creating facts, fact tables, dimension attributes, or dimensions is contextual. The principle of the $M2^{*}_{Rules}$ is thus that of semi automatic rules that help identifying possible candidate for a mapping but leave the actual decision of the mapping to the designer.

The Star schema was chosen to structure the target multidimensional model (Kimball, 1995) because it is widely, if not the most, used in industry. The Star model should respect a number of design constraints. CADWA proposes to normalise the draft model using *Star* heuristic rules ($*_{Rules}$). So far, 10 $*_{Rules}$ and 15 $M2^{*}_{Rules}$ have been defined. For the sake of space only 5 examples of $*_{Rules}$ and 8 $M2^{*}_{Rules}$, noted SR1 to SR5 and M2SR1 to M2SR8, are presented below.

SR1: In the dimensional models, the facts tables express the relation of one to many between the dimension tables

SR2: A line in a fact table corresponds to several measures. A measure is an attribute in a fact table. All the measures of a same fact table of facts must have the same granularity

SR3: A fact can be numeric additive, semi-additive (can be added only for certain dimensions) or non-additive (can't be added). The most useful facts of a fact table are numeric and additive.

SR4: A table of dimension contains several attributes.

SR5: Attributes of dimensions allow varying the possibilities of analyses in slices and dices.

M2SR1: The parameter "destination" of direction is a dimension table or dimension attribute.

M2SR2: The parameter "object" of the target is a fact

M2SR3: The parameter "Result" of the target is a fact

M2SR4: The parameter “Result” of the target is a fact and dimension table according to a particular context

M2SR5: The parameter source of a direction is dimension table

M2SR6: A location is a dimension table

M2SR7: a beneficiary is a dimension table

M2SR8: An actor is a dimension table

In the BR example, applying $M2^{*Rules}$ to the generalised version of AI1 “Sell (verb) a quantity (quantity) of products (object) in stores (location) during certain dates (time)” provides two modelling indications: (i) three dimensions are expected in the DW model: product, store, and date, and (ii) a fact can be created: sales quantity.

The star model shown in Fig. 4 presents a piece of the DW model produced by applying $*Rules$ to the model fragment produced with the two aforementioned modelling indications (cf. “extract modelling indications” activity). The model shown in Fig. 5 is produced based on the modelling indications discovered from all action intentions of the Marketing Manager and extracting additional knowledge from reusable data marts as explained in the next section.

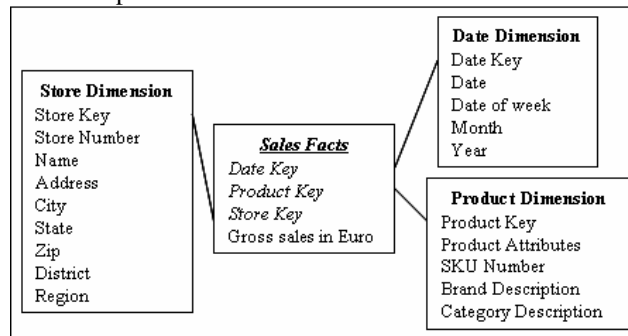


Fig. 4: Model chunk produced by applying $M2^{*Rules}$ and $*Rules$ to AI1.

3.2 Reuse and Selection of DM models

A number of packaged solutions exist in the market to facilitate DW development and exploitation. *CADWA* takes into account the existence of activity-oriented reusable DM by reusing their models and adapting them to the context of the organization.

CADWA proposes a repository of reusable DM model fragments. Each model fragment can be exploited to support one or several measures in a given activity sector (cf. “Select reusable DM” and “propose DwMF” activities). The *CADWA* repository can be extended by new models, either by accumulating experience with packaged solutions, or by generalising solutions already defined in a specific organizational context so as to use them in other contexts (cf. “Adapt DM” activity).

For example, generic DM models for CRM and sales modules of Business Object, Micro Strategy, Oracle, SAS Hyperion, Informatica or Information builders are recorded in the *CADWA* repository. They are reused to identify the standard indicators that the Marketing Manager is likely to look when making decisions, and therefore to identify the required fact tables and dimensions in the developed DW model. The dimensions and fact tables shown in Fig. 5 are enriched using the repository of

reusable DM (cf. “generate new DW” activity). The table attributes have been added for their ability to help record the data needed in evaluating indicators that can be considered to monitor the Marketing Manager MiBP.

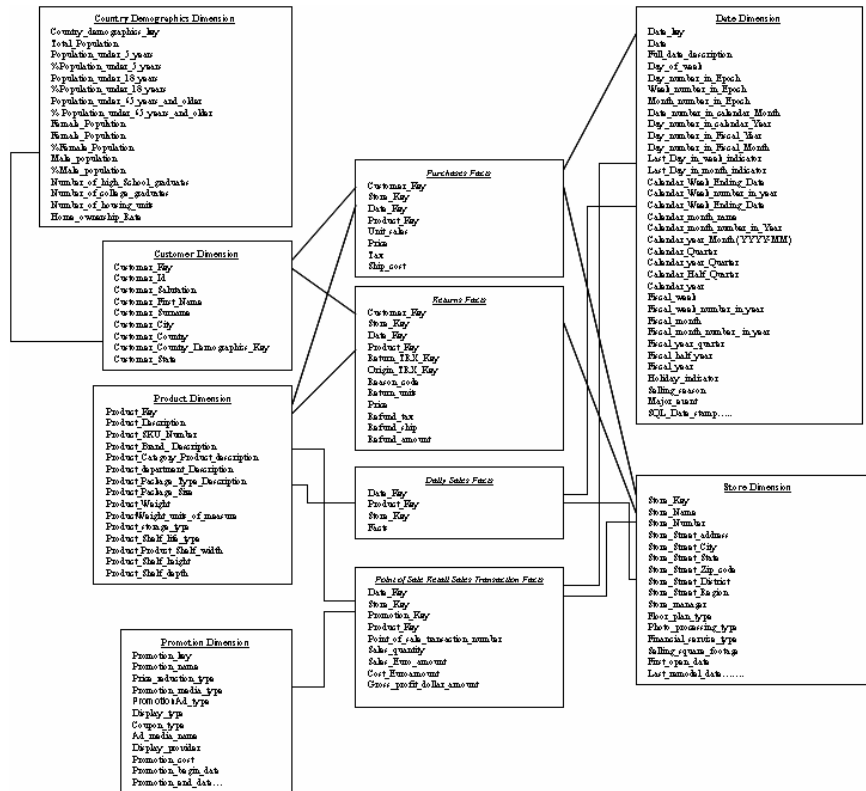


Fig. 5. DW Model

4. Integration of models

Other development constraints than the user requirements considered so far should be taken in consideration in the DW data model design. For example, efficiency and response time, system availability, data quality, flexibility, or consistency with legacy systems.

At the design level, the DW model can be modified while integrating it with the *legacy DW* (cf. “abstract legacy DW” activity) and taking into consideration the *existing data models* (cf. “abstract DB” activity) of the operational data sources. The other constraints can be defined while constructing the physical database. Other research approaches and industry tools (Vassiliadis and al. 1999) can be used to deal with these issues that we consider outside the scope of CADWA.

5. Related works

A few proposals have been made to deal with DW requirements elicitation and analysis by establishing the relationship between the DW and its organizational context of use.

(Prakash and al., 2004) propose a requirement elicitation process for data warehouses that groups requirements in several levels of abstraction. Their approach is to identify the DW requirements via information scenarios. This process exploits a Goal-Decision-Information (GDI) diagram. It starts with the determination of the organization goals. Then, the decisions that need to be made are identified. Finally, the information needed to make these decisions is specified. Although, they show how to obtain the GDI diagram and the information scenarios, the authors do not articulate the relationships between information scenarios and requirements.

(Schiefer et al., 2002) present a method, easyREMOTEDWH (easy Requirements Modeling Technique for Data Warehouses) that considers DW requirements from different stakeholders' perspectives. Their approach is also based on goal modelling at several levels of abstraction. They propose an interesting approach to represent business objectives and needs. Unfortunately, they do not propose any notation or guidelines to properly specify the DW requirements, or to guide their analysis in the later phase of the DW design process.

(Böhnlein et al., 2000) propose a different approach that starts from business process models to elicit the DW requirements. This approach is similar to the data driven approach, as it focuses on the subject of the monitoring, rather than on the reason for which monitoring is required. In other terms, it is opportunistic rather than really driven by requirements.

An alternative approach is proposed by Winter and Straüch [WS03] [WS04] who suggest to focus on the decision processes themselves. However, their study points out that a detailed business process analysis is not always feasible because decision processes consist of unique and unstructured tasks. Besides, decision makers often refuse to disclose their processes.

(Paim et al., 2003) present the DWARF (Data Warehouse Requirements deFinition) technique. They adapt a traditional requirements engineering method (the NFR Framework) to the specific issues raised by data warehouses. Their approach is able to capture and deal with non-functional requirements on the technical level (e.g. how to access data), but it is not able to deal with the adequacy of the DW model for the decisions that will be made using it. Similarly, the Goal-Question-Metric approach is adapted by (Vassiliadis and al. 1999) to take NFRs into account in the physical architecture of the DW.

6. Conclusion

We believe that DW that better suit the decision-makers expectations would be delivered more easily by better focusing on their requirements in the initial design phase. This paper presents the *CADWA* method that proposes to take into account decision-makers' requirements with a goal oriented approach. The guidance offered

by *CADWA* helps eliciting early requirements with organization business plan, refining them into Decision-maker macro and micro business plans, and concretizing the corresponding action plans into a data DW model.

A first evaluation of *CADWA* was achieved by interviewing two DW development experts⁵. The questions raised during the interviews aimed at testing the scalability of the approach, the quality of the help that it is likely to bring, and its easiness of use. Several practical issues were raised:

The repository of reusable models should be indexed by domain of activity. The question of where such models come from was asked. In the context of globalisation, developing DW compatible with standard models is an advantage for improving flexibility, and easiness of use for new users. SME also see the reuse of standard DW models as a way to improve their way to make decisions. However, our experts raised our attention on the fact that very few consultants are likely to spend the effort and time generalising and documenting the knowledge acquired through years. An approach must thus be proposed to facilitate this activity. Providing built-in standard models in *CADWA* would be an even better convincing argument for a technology transfer. We, however, have the feeling that this is outside the scope of our research. Another discussion thread concerned the complexity of organization structures. The idea of dispatching goals between decision makers to separate concerns was considered as realistic and useful, even though it shall raise political issues in large organizations. However, our experts noticed that nothing was proposed in *CADWA* to guide goal affectation, or the impact analysis of goal changes in the context of complex organizations. In practice, the question is how to take a new high level enterprise goal into consideration at the level of subsidiaries, and what kind of analysis should systematically be made to understand the impact of this change on the DW model.

Last, our experts underlined that, in their experience, the time available for gathering requirements with decision makers is about 10 to 15 two-hour meetings for a 2-years DW project. This is of course very short, considering the dimension of DW projects. An effort should thus be made to better understand which part of the method requires presence of the decision maker, and which does not. Alternative approaches for accelerating requirements identification should also be proposed.

We are in the process of formalizing the *CADWA* method in order to prepare the development of a support tool. Some of the comments obtained during our first evaluation were already taken into account, but as shown the interview report, more efforts are needed to reach the target level of quality. We plan to proceed with evaluation by developing a questionnaire on DW development practice, and a real case study provided by our industrial contacts.

7. References

Ben Achour C., Ncube C., "Engineering the PORE Method for COTS Selection and Implementation with the MAP Process Meta-Model", *Proceedings of REFSQ'2000*, 6th

⁵ Acknowledgement: we would like to thank Mr Cedric Baret and Mr Nessim Toumi for their useful remarks in order to ameliorate our *CADWA* approach.

- International Workshop on Requirements Engineering Foundations of Software Quality*, Stockholm, Sweden, 2000.
- Boehnlein M., Vom Ende U., "A Business Process Oriented Development of Data Warehouse Structures". In *Proceedings of Data Warehousing 2000*, Physica Verlag, 2000.
- Freudi M., Salinesi C., "Requirement engineering for data warehousing" *Proc. of the 9th International Workshop on Requirements Engineering : Foundations of Software Quality (REFSQ'03)*, 2003.
- Kimball R., "*The Data Warehouse toolkit*". John Wiley & Sons, 1996.
- Mazón J.N., Trujillo J., Serrano M., Piattini M., "Designing Data Warehouses: From Business Requirement Analysis to Multidimensional Modeling". *REBNITA '05*, Paris, France, 2005.
- Paim F.R., Castro J.B., "DWARF: An Approach for Requirements Definition and Management of Data Warehouse System". *11th IEEE International Requirements Engineering Conference (RE'03)*. Monterey Bay, California, USA, 2003.
- Prakash N., Gosain A., "A Requirements Engineering for Data warehouse Development". *Proceedings of CAiSE03 Forum*, 2003.
- Prakash N., Singh Y., Gosain A., "Information scenarios for data warehouse requirements elicitation". *23th International Conference on Conceptual Modelling (ER2004)*. Volume 3288 of LNCS, pages 205-216. Springer-Verlag (2004).
- Prat N., "Goal formalisation and classification for requirements engineering", *Proc. of the 3rd International Workshop on Requirements Engineering : Foundations of Software Quality (REFSQ'97)*, 1997.
- Rilson F., Freire J., "DWARF: An Approach for Requirements Definition and Management of Data Warehouse Systems". *Proceeding of the 11th IEEE International Requirements Engineering Conference*, September 08 - 12 (2003), pages 1090-1099.
- Salinesi C., Rolland C., "Fitting Business Models to Systems Functionality Exploring the Fitness Relationship". *Proceedings of the 15th Conference on Advanced Information Systems Engineering (CAiSE'03)*, Klagenfurt/Velden, Austria, 2003.
- Schiefer J., List B., Bruckner R., "A holistic approach for managing requirements of data warehouse systems". In *Proceeding of Americas Conference on Information Systems*, 2002.
- Vassiliadis P., Bouzeghoub M., Quix, C., "Toward Quality-Oriented Data warehouse Usage and Evolution", *11th Internat. Conf. On Advanced Information Systems Engineering (CAiSE'99)*, Lecture Notes in Computer Science, Springer, Heidelberg, June 1999.
- Winter R., Strauch B., "A method for demand-driven Information Requirements Analysis in Data Warehousing Projects". *Proceeding of the 36th Hawaii International Conference on System Sciences*, 2003.
- Winter R., Strauch B., "Information Requirements Engineering for Data Warehouse Systems". *ACM Symposium on Applied Computing (SAC'04) Nicosia, Cyprus*, 2004.